# Multi-Frequency Phase Synchronization

Tingran Gao

William H. Kruskal Instructor
Committee on Computational and Applied Mathematics (CCAM)
Department of Statistics
The University of Chicago

Applied Algebra Seminar
University of Wisconsin–Madison
Madison, WI

April 26, 2019

# Outline

**Motivation**

- Synchronization
- Geometric Model of Fibre Bundles

Multi-Frequency Phase Synchronization

- Phase Synchronization
- Multi-Frequency Formulation
- Theory

Application to Multi-Frequency Class Averaging

Joint work with **Zhizhen Zhao (UIUC)**

# Outline

**Motivation**

- **Synchronization**
- Geometric Model of Fibre Bundles

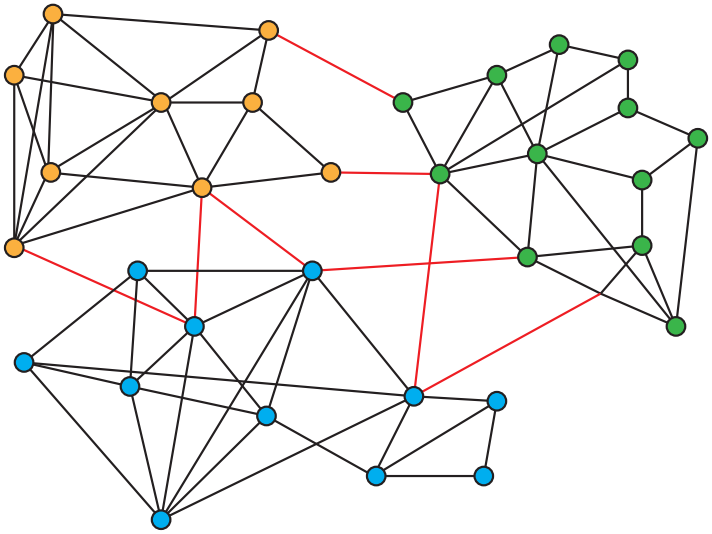
Multi-Frequency Phase Synchronization

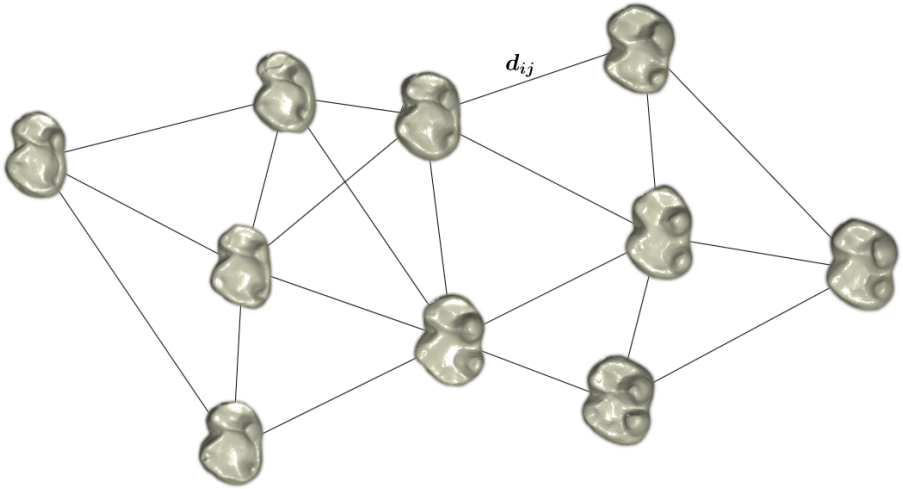
- Phase Synchronization
- Multi-Frequency Formulation
- Theory

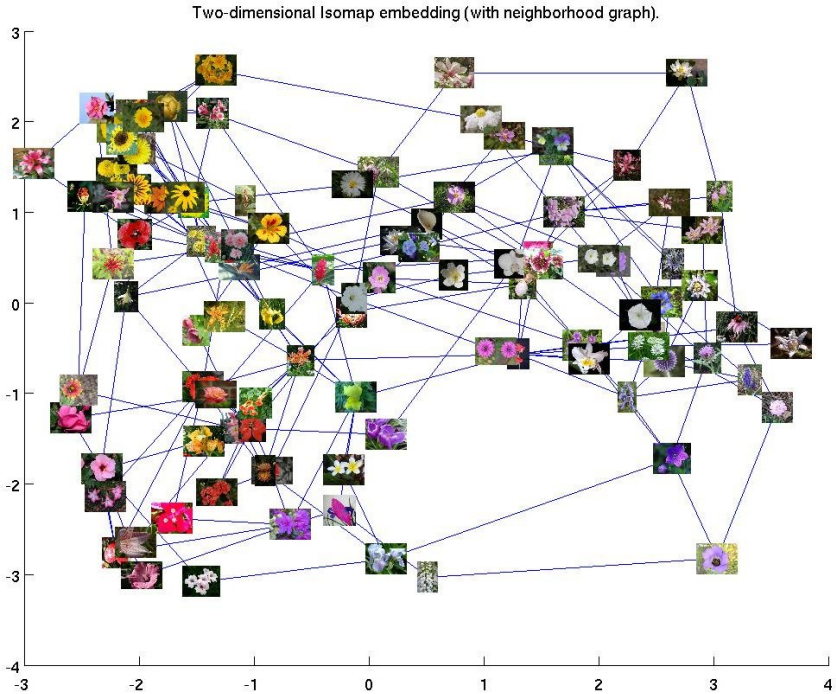
Application to Multi-Frequency Class Averaging

Joint work with **Zhizhen Zhao (UIUC)**

# Graph-Based Data Analysis



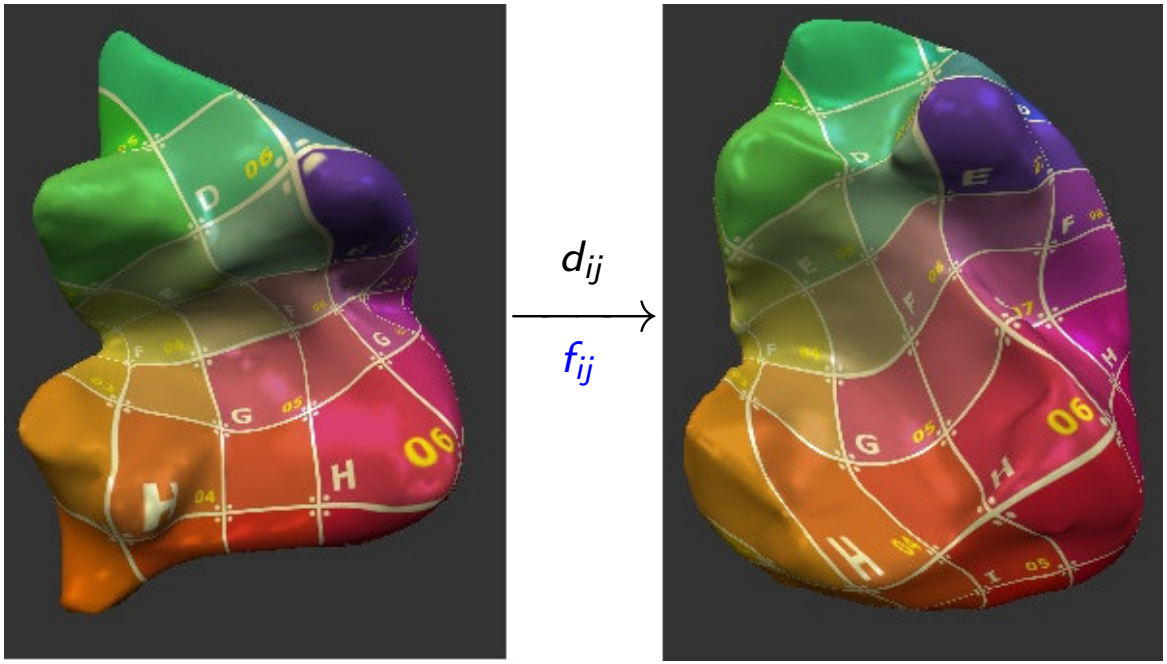


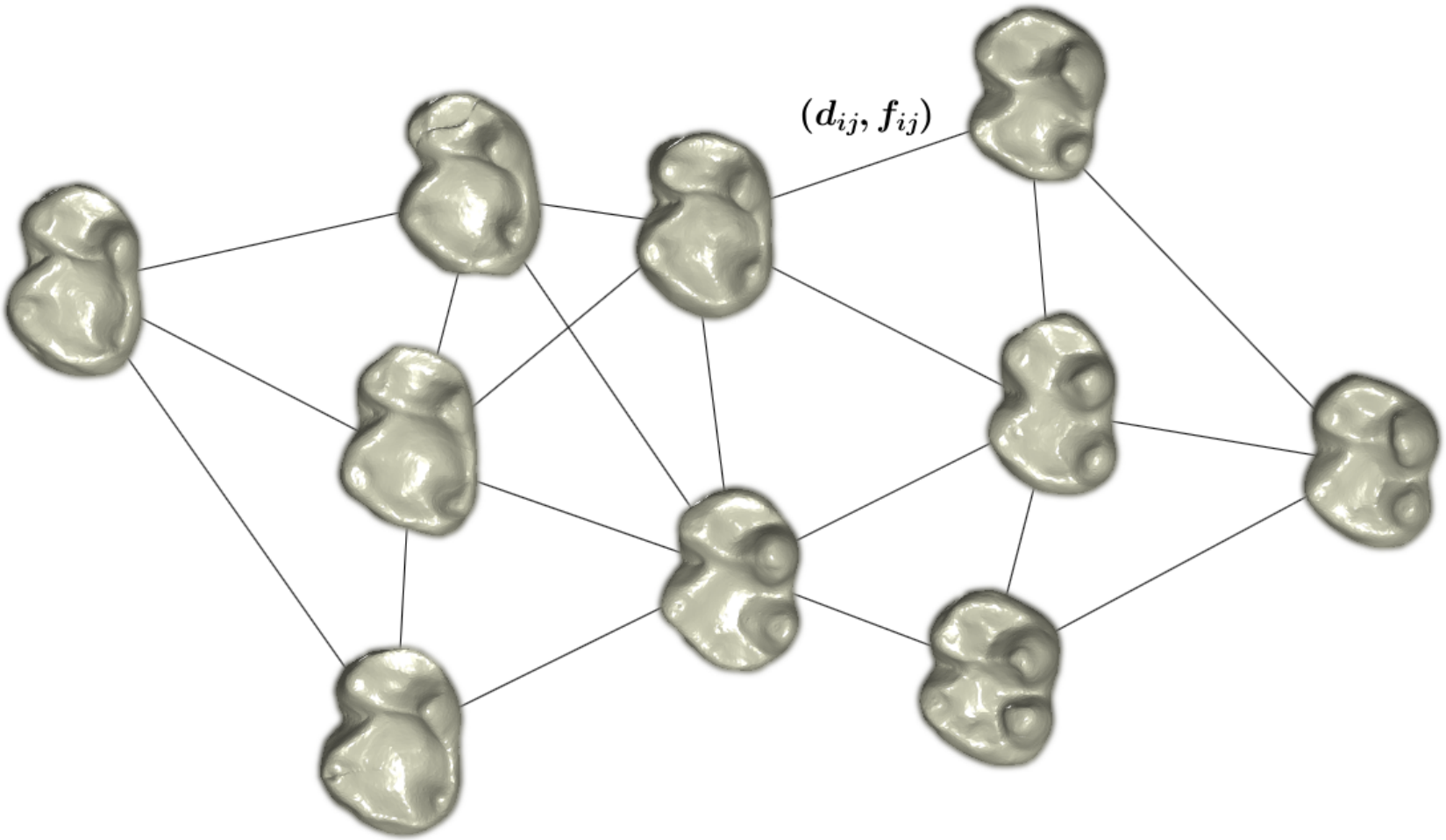
# Evolutionary Biology

$$d_{\mathrm{cP}}\left(S_i, S_j\right) = \inf_{\mathcal{C}\in\mathcal{A}\left(S_i,S_j\right)} \inf_{R\in\mathbb{E}(3)} \left(\int_{S_i} \left\|R\left(x\right) - \mathcal{C}\left(x\right)\right\|^2 d\mathrm{vol}_{S_i}\left(x\right)\right)^{\frac{1}{2}}$$

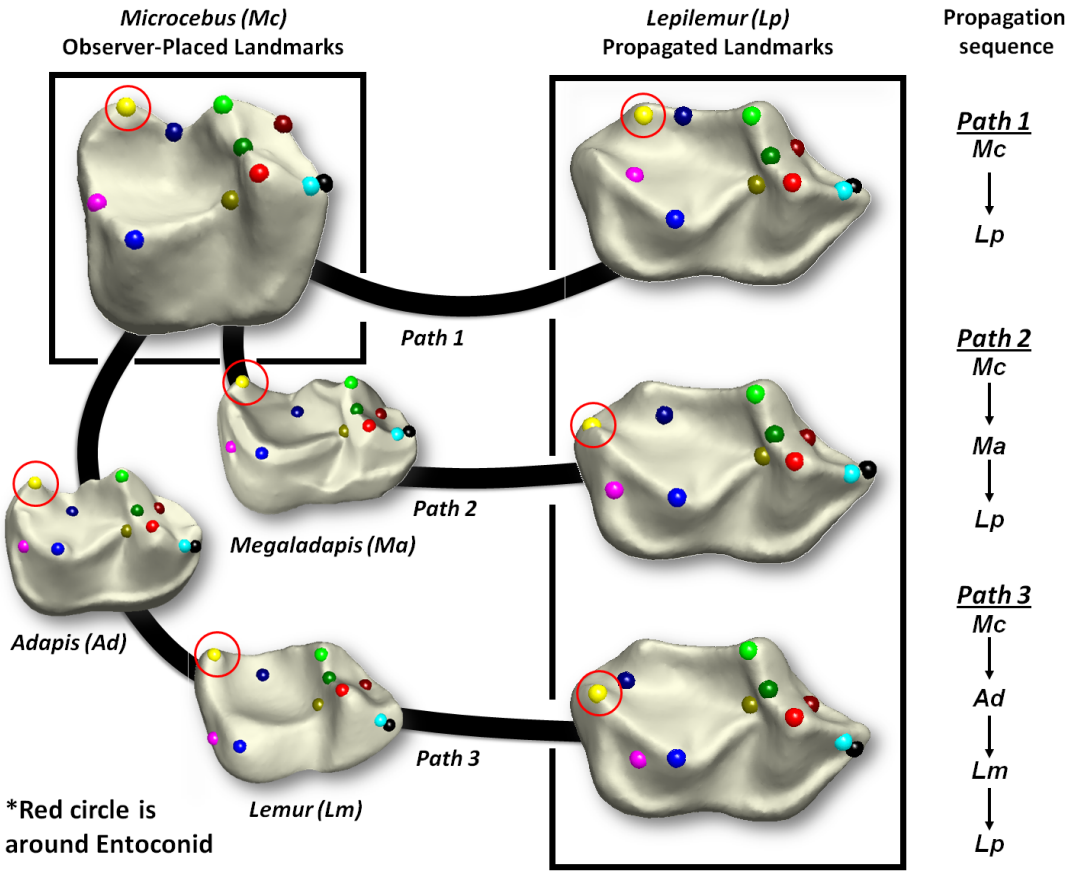
$S_i$




$S_j$

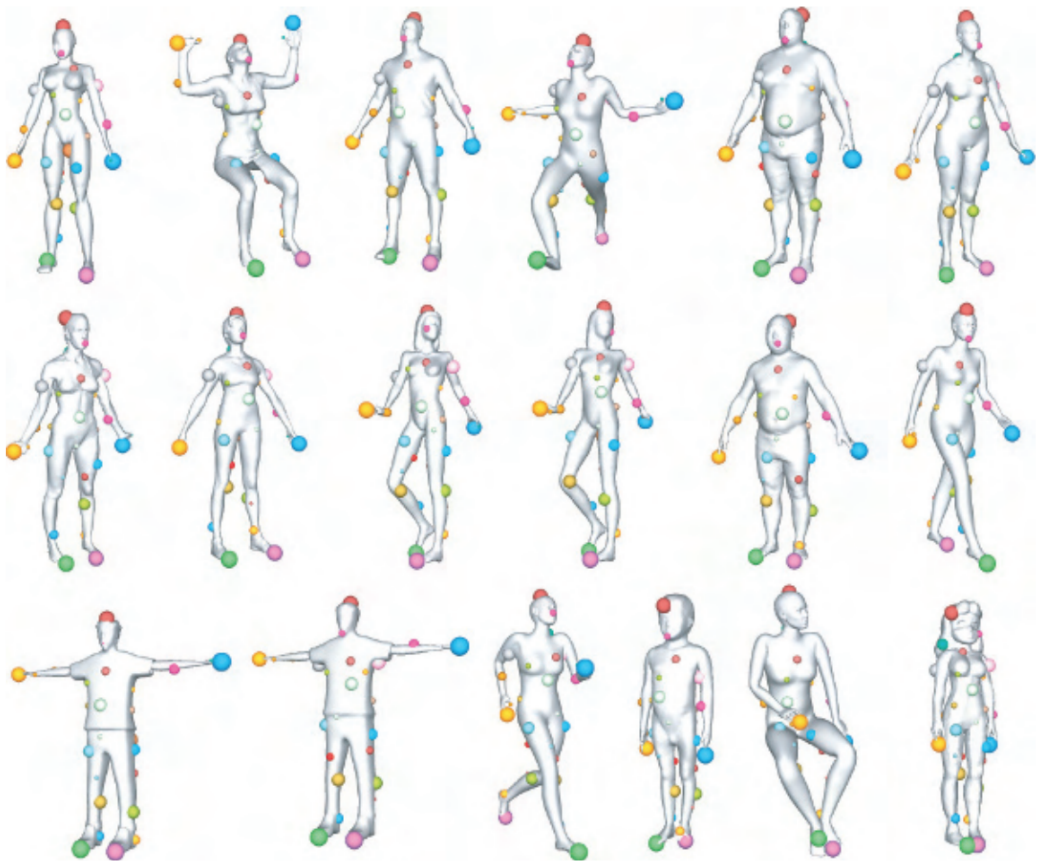

$(d_{ij}, f_{ij})$

# Interpretability Issue



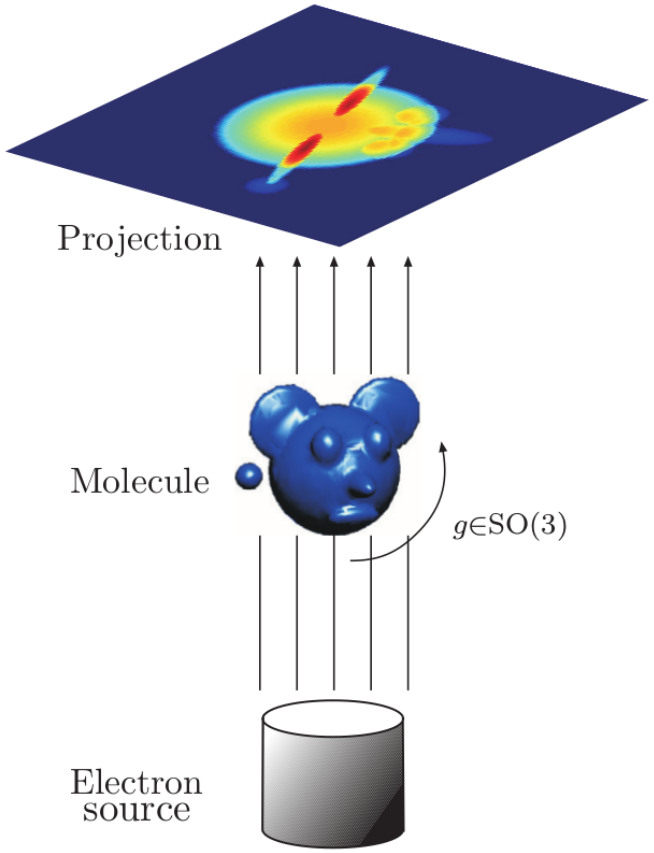

• Boyer et al. "Algorithms to Automatically Quantify the Geometric Similarity of Anatomical Surfaces." *Proceedings of the National Academy of Sciences* 108.45 (2011): 18221–18226.

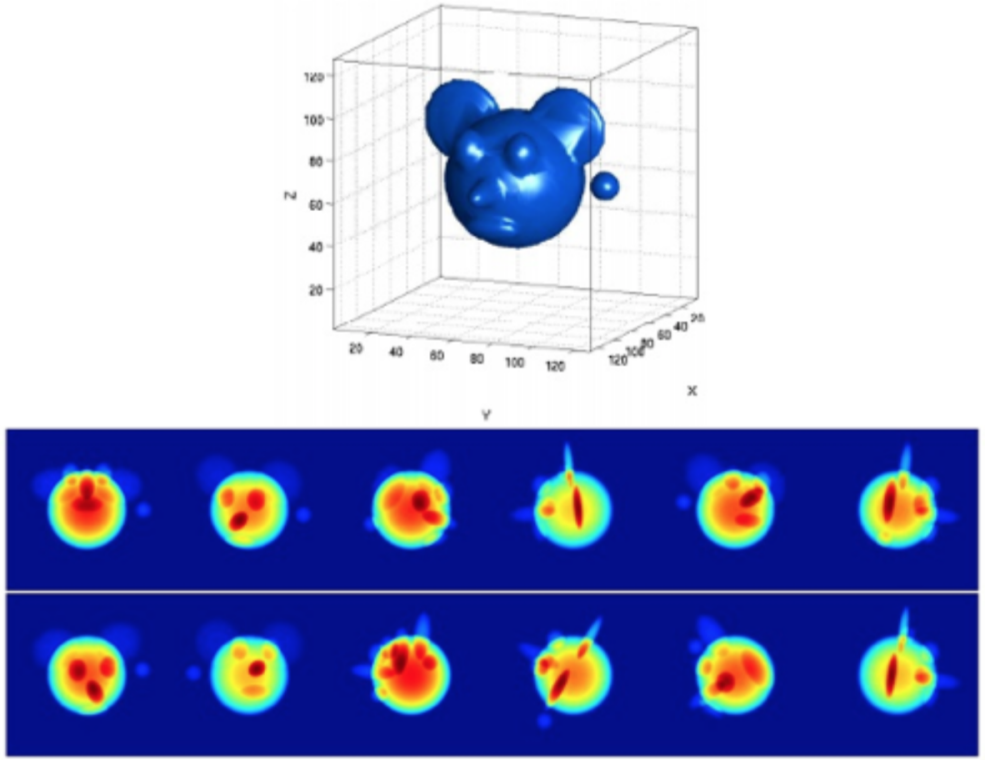
# Collection Shape Analysis



• Q. Huang & L. Guibas. "Consistent Shape Maps via Semidefinite Programming", *Computer Graphics Forum*, vol. 32, no. 5, pp. 177–186 (2013).

# Cryo-Electron Microscopy



• Singer et al. "Viewing Angle Classification of Cryo-Electron Microscopy Images using Eigenvectors", *SIAM Journal on Imaging Sciences*, 4 (2), pp. 543–572 (2011).

# Cryo-Electron Microscopy: Real Challenge is Low SNR

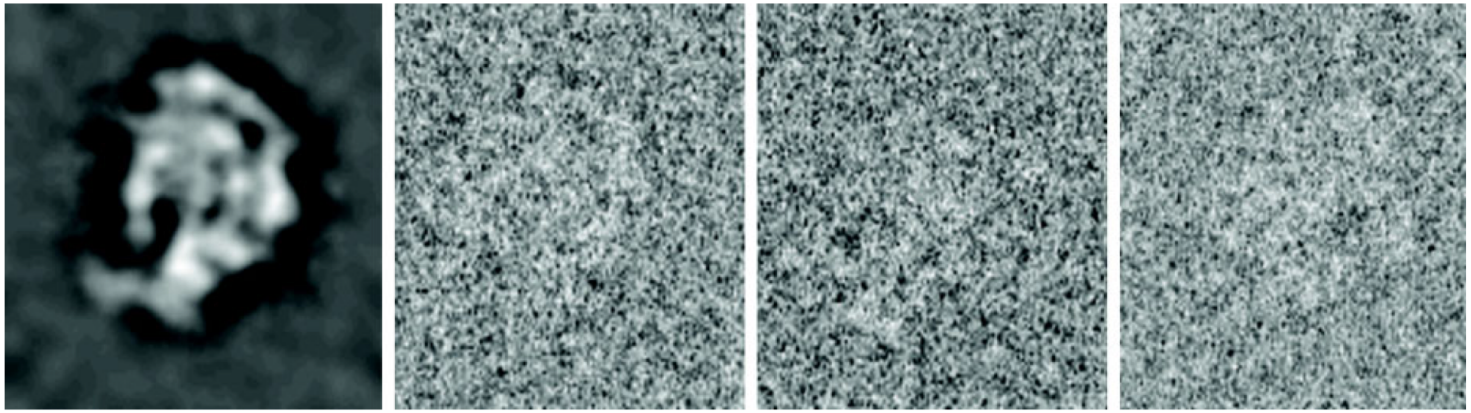
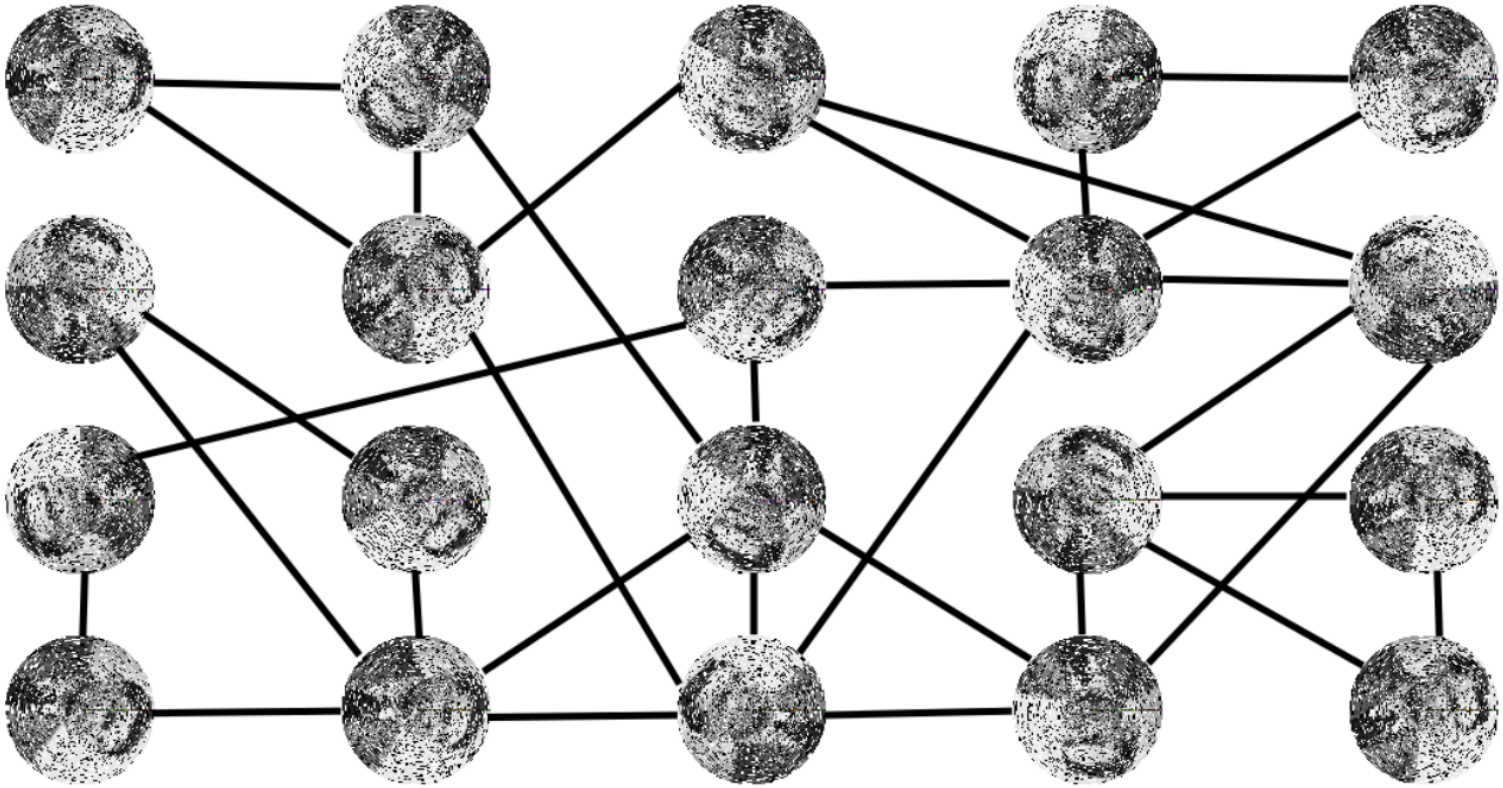


**Fig. 3** The left most image is a clean simulated projection image of the E.coli 50S ribosomal subunit. The other three images are real electron microscope images of the same subunit

• Hadani & Singer. "Representation Theoretic Patterns in Three-Dimensional Cryo-Electron Microscopy II – The Class Averaging Problem," *Foundations of Computational Mathematics*, 11 (5), pp. 589–616 (2011).

$$y_i = R_i x + \xi_i$$
$$R_i \in O(d), \quad \xi_i \sim \text{i.i.d. noise}$$



- Afonso S. Bandeira. "Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science." (2015).

$$\boxed{\begin{array}{c} y_i = R_i x + \xi_i \\ R_i \in O(d), \quad \xi_i \sim \text{i.i.d. noise} \end{array}}$$

$$\begin{aligned} \text{Measurement:} \quad & R_{ij} \approx R_i^\top R_j \\ \text{Recover:} \quad & R_1, R_2, \cdots \end{aligned}$$

$$\boxed{\begin{array}{c} y_i = R_i x + \xi_i \\ R_i \in O(d), \quad \xi_i \sim \text{i.i.d. noise} \end{array}}$$

$$\begin{aligned} \text{Measurement:} \quad & R_{ij} \approx R_i^\top R_j \\ \text{Recover:} \quad & R_1, R_2, \cdots \end{aligned}$$

$\Rightarrow$ Solve the minimization problem

$$\min_{R_1, R_2, \cdots \in O(d)} \quad \sum_{(i,j) \in E} w_{ij} \left\| R_{ij} - R_i^\top R_j \right\|_{\mathrm{F}}^2$$

$$\frac{1}{2} \sum_{1 \leq i,j \leq N} w_{ij} \left\| R_{ij} - R_i^\top R_j \right\|_{\mathrm{F}}^2$$

$$
\begin{aligned}
&\frac{1}{2}\sum_{1\le i,j\le N} w_{ij}\left\|R_{ij}-R_i^\top R_j\right\|_{\mathrm{F}}^2 \\
&\quad=\frac{1}{2}\sum_{1\le i,j\le N} w_{ij}\mathrm{Tr}\left[\left(R_i^\top-R_{ij}R_j^\top\right)\left(R_i-R_jR_{ij}^\top\right)\right] \\
&\quad=\sum_{1\le i,j\le N} w_{ij}\mathrm{Tr}\left(I_{d\times d}-R_iR_{ij}R_j^\top\right)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{2}\sum_{1\le i,j\le N} w_{ij}\left\|R_{ij}-R_i^\top R_j\right\|_{\mathrm{F}}^2 \\
&\quad=\frac{1}{2}\sum_{1\le i,j\le N} w_{ij}\mathrm{Tr}\left[\left(R_i^\top-R_{ij}R_j^\top\right)\left(R_i-R_jR_{ij}^\top\right)\right] \\
&\quad=\sum_{1\le i,j\le N} w_{ij}\mathrm{Tr}\left(I_{d\times d}-R_iR_{ij}R_j^\top\right) \\
&\quad=\mathrm{Tr}\left(\left[R_1,\cdots,R_N\right]\left(\mathscr{D}-\mathscr{W}\right)\begin{bmatrix}R_1^\top\\ \vdots\\ R_N^\top\end{bmatrix}\right)
\end{aligned}
$$

where

$$
\mathscr{D}=\begin{bmatrix} d_1 I_{d\times d} & & \\ & \ddots & \\ & & d_N I_{d\times d}\end{bmatrix},\mathscr{W}=\begin{bmatrix} w_{11}R_{11} & \cdots & w_{1N}R_{1N}\\ \vdots & & \vdots\\ w_{N1}R_{N1} & \cdots & w_{NN}R_{NN}\end{bmatrix}
$$

# QO-OC

$$\min_{R_1,R_2,\cdots\in O(d)} \quad \sum_{(i,j)\in E} w_{ij} \left\| R_{ij} - R_i^\top R_j \right\|_{\mathrm{F}}^2$$

$$\Leftrightarrow \quad \min_{R_1,\cdots,R_N\in\mathbb{R}^{d\times d}} \mathrm{Tr}\left( \left[R_1,\cdots,R_N\right] \left(\mathscr{D}-\mathscr{W}\right) \begin{bmatrix} R_1^\top \\ \vdots \\ R_N^\top \end{bmatrix} \right)$$

$$\text{s.t.} \qquad R_1^\top R_1 = \cdots = R_N^\top R_N = I_{d\times d}$$

• A. Nemirovski, "Sums of Random Symmetric Matrices and Quadratic Optimization Under Orthogonality Constraints", *Math. Programming*, 109 (2007), pp. 283–317.

# Convex Relaxations (I)

$$\min_{R_1,\cdots,R_N \in \mathbb{R}^{d\times d}} \operatorname{Tr}\left( \left[R_1, \cdots, R_N\right] \left(\mathscr{D} - \mathscr{W}\right) \begin{bmatrix} R_1^\top \\ \vdots \\ R_N^\top \end{bmatrix} \right)$$
$$\text{s.t.} \qquad R_1^\top R_1 = \cdots = R_N^\top R_N = I_{d\times d}$$

- Spectral Relaxation

$$\boxed{\begin{aligned} &\min_{\mathcal{R} \in \mathbb{R}^{d\times Nd}} \operatorname{Tr}\left(\mathcal{R}\left(\mathscr{D} - \mathscr{W}\right)\mathcal{R}^\top\right) \\ &\text{s.t.} \qquad \mathcal{R}\mathcal{R}^\top = N \cdot I_{d\times d} \end{aligned}}$$

• Chaudhury et al. "Global Registration of Multiple Point Clouds using Semidefinite Programming", *SIAM Journal on Optimization*, 25(1):468–501, 2015.

# Convex Relaxations (II)

$$\min_{R_1,\cdots,R_N \in \mathbb{R}^{d\times d}} \operatorname{Tr}\left( (\mathscr{D} - \mathscr{W}) \begin{bmatrix} R_1^\top \\ \vdots \\ R_N^\top \end{bmatrix} \begin{bmatrix} R_1, \cdots, R_N \end{bmatrix} \right)$$
$$\text{s.t.} \qquad R_1^\top R_1 = \cdots = R_N^\top R_N = I_{d\times d}$$

- 

$$\boxed{\begin{aligned} \min_{\mathcal{G} \in \mathbb{R}^{Nd\times Nd}} \quad & \operatorname{Tr}\left( (\mathscr{D} - \mathscr{W})\, \mathcal{G} \right) \\ \text{s.t.} \quad & \mathcal{G} \succeq 0,\ \mathcal{G} = \mathcal{G}^\top \\ & \mathcal{G}_{ii} = I_{d\times d},\ 1 \le i \le N \\ & \operatorname{rank}(\mathcal{G}) = d \end{aligned}}$$

• Chaudhury et al. "Global Registration of Multiple Point Clouds using Semidefinite Programming", *SIAM Journal on Optimization*, 25(1):468–501, 2015.

# Convex Relaxations (II)

$$\min_{R_1,\cdots,R_N \in \mathbb{R}^{d\times d}} \operatorname{Tr}\left( (\mathscr{D} - \mathscr{W}) \begin{bmatrix} R_1^\top \\ \vdots \\ R_N^\top \end{bmatrix} \begin{bmatrix} R_1, \cdots, R_N \end{bmatrix} \right)$$
$$\text{s.t.} \qquad R_1^\top R_1 = \cdots = R_N^\top R_N = I_{d\times d}$$

- SDP Relaxation

$$\boxed{\begin{aligned} \min_{\mathcal{G} \in \mathbb{R}^{Nd\times Nd}} \quad & \operatorname{Tr}\left( (\mathscr{D} - \mathscr{W})\, \mathcal{G} \right) \\ \text{s.t.} \quad & \mathcal{G} \succeq 0,\ \mathcal{G} = \mathcal{G}^\top \\ & \mathcal{G}_{ii} = I_{d\times d},\ 1 \le i \le N \\ & \cancel{\operatorname{rank}(\mathcal{G}) = d} \end{aligned}}$$

• Chaudhury et al. "Global Registration of Multiple Point Clouds using Semidefinite Programming", *SIAM Journal on Optimization*, 25(1):468–501, 2015.

# Outline

**Motivation**

- Synchronization
- **Geometric Model of Fibre Bundles**

Multi-Frequency Phase Synchronization

- Phase Synchronization
- Multi-Frequency Formulation
- Theory

Application to Multi-Frequency Class Averaging

Joint work with **Zhizhen Zhao (UIUC)**

# Synchronization Problems

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $\rho : E \to G$ satisfying $\rho_{ij} = \rho_{ji}^{-1}$, $\forall (i,j) \in E$
- Goal:
  - find a **vertex potential** $f : V \to G$ or $F$ such that

$$f_i = \rho_{ij} f_j, \quad \forall (i,j) \in E$$

- The goal can be achieved if and only if $\rho_{ij} = f_i f_j^{-1}$
- **Not** always feasible!

# Synchronization Problems

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $\rho : E \to G$ satisfying $\rho_{ij} = \rho_{ji}^{-1}$, $\forall (i,j) \in E$
- Goal:
  - find a **vertex potential** $f : V \to G$ or $F$ such that

$$f_i = \rho_{ij} f_j, \quad \forall (i,j) \in E$$

- The goal can be achieved if and only if $\rho_{ij} = f_i f_j^{-1}$
- **Not** always feasible!
- If infeasible, find the "closest solution" in the sense of

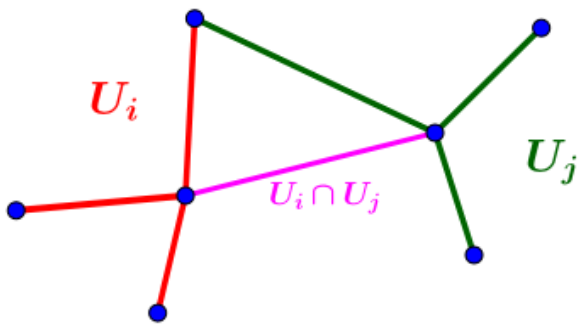
$$\min_{\substack{f:V\to G \\ \|f\|\neq 0}} \frac{1}{2} \frac{\sum_{i,j\in V} \|f_i - \rho_{ij} f_j\|^2}{\sum_{i\in V} \|f_i\|^2} (=: \eta(f))$$

# Synchronization Problems: Examples

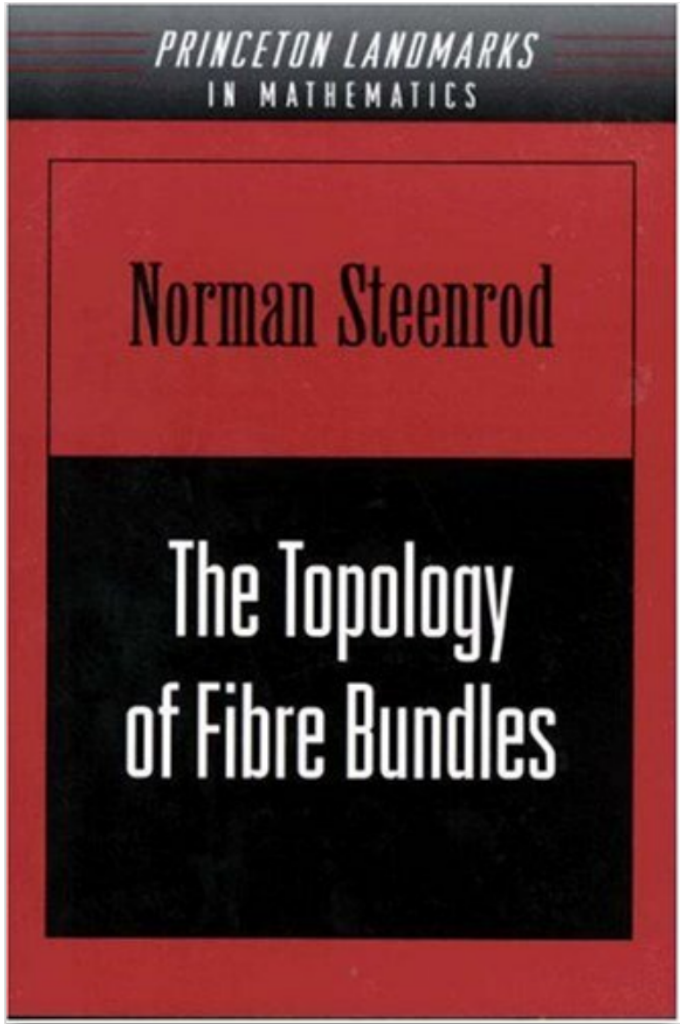
- Manifold Orientability [Singer, Wu (2011)]: $G = O(1)$
- Angular Synchronization [Singer (2011)]: $G = U(1)$
- Vector Diffusion Maps [Singer, Wu (2012)]: $G = O(d)$
- Multireference Alignment [Bandeira et al. (2014)]: $G = \{\text{cyclic shifts}\}$
- Global Registration of Point Clouds [Chaudhury (2015)]: $G = \mathbb{E}_d$
- Collection Shape Matching [Nguyen et al. (2011)], [Huang, Guibas (2013)], [Chen et al. (2014)], [Maron et al. (2016)]: $G = \text{permutation group}$
- Cryo-EM Structural Reconstruction [Singer et al. (2011)], [Shkolnisky, Singer (2012)], [Zhao, Singer (2014)], [Bandeira et al. (2015)]: $G = SO(3)$

# The Geometry of Synchronization Problems

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $\rho : E \to G$ satisfying $\rho_{ij} = \rho_{ji}^{-1}$, $\forall (i,j) \in E$
- Flat Principal Bundle:
  - Let $\mathfrak{U} = \{U_i \mid 1 \leq i \leq |V|\}$ be an open cover of $\Gamma$ (viewed as a 1-dimensional simplicial complex), where $U_i$ is the *(open) star neighborhood* of vertex $i$.



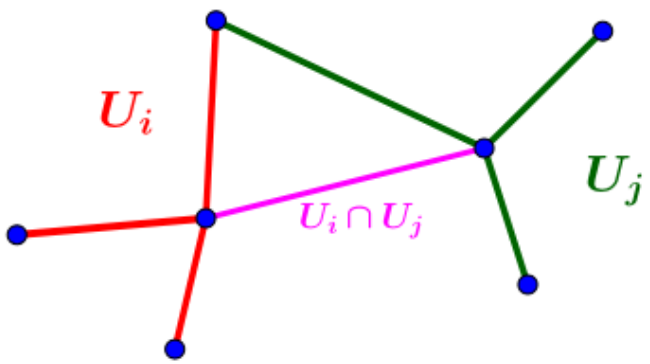

  - Triplet $(\rho, G, \Gamma)$ defines a *flat principal $G$-bundle* $\mathscr{B}_\rho$ over $\Gamma$




**Theorem** (Steenrod 1951, §2). If topological group $G$ acts on $Y$ and $\{U_i\}$, $\{g_{ij}\}$ is a system of coordinate transformations in the space $X$ such that

$$\begin{aligned} g_{ii} &= e \in G \quad \text{for all } U_i \\ g_{ij} &= g_{ji}^{-1} \quad \text{if } U_i \cap U_j \neq \emptyset \\ g_{ij} g_{jk} &= g_{ik} \quad \text{if } U_i \cap U_j \cap U_k \neq \emptyset \end{aligned}$$

then there exists a fibre bundle $\mathscr{B}$ with base space $X$, fibre $Y$, group $G$, and coordinate transforms $\{g_{ij}\}$.




**No triple intersections!**

**Theorem** (Steenrod 1951, §2). If topological group $G$ acts on $Y$ and $\{U_i\}$, $\{g_{ij}\}$ is a system of coordinate transformations in the space $X$ such that

$$g_{ii} = e \in G \quad \text{for all } U_i$$

$$g_{ij} = g_{ji}^{-1} \quad \text{if } U_i \cap U_j \neq \emptyset$$

$$\cancel{g_{ij}g_{jk} = g_{ik} \quad \text{if } U_i \cap U_j \cap U_k \neq \emptyset}$$

then there exists a fibre bundle $\mathscr{B}$ with base space $X$, fibre $Y$, group $G$, and coordinate transforms $\{g_{ij}\}$.

# Geometric Observations

- Denote

$$C^0(\Gamma; G) := \{f : V \to G\}$$
$$C^1(\Gamma; G) := \left\{\rho : E \to G \mid \rho_{ij} = \rho_{ji}^{-1}, \forall (i,j) \in E\right\}$$

- Consider the right action of $C^0(\Gamma; G)$ on $C^1(\Gamma; G)$:

$$\begin{aligned} C^1(\Gamma; G) \times C^0(\Gamma; G) &\to C^1(\Gamma; G) \\ (\rho, f) &\longmapsto \tau_\rho f \end{aligned}$$

  defined as $(\tau_f \rho)_{ij} := f_i^{-1} \rho_{ij} f_j, \quad \forall (i,j) \in E$. Denote the orbit space of $\tau_f$ by $C^1(\Gamma; G) / C^0(\Gamma; G)$

- $\rho$ synchronizable $\Leftrightarrow$ $\tau_f \rho$ synchronizable for all $f \in C^0(\Gamma; G)$, i.e. synchronizability is defined up to orbits of $\tau_f$

# Moduli Space of Synchronization Data

**Theorem (G. et al. 2019).** There exists a one-to-one correspondence (between *sets*)

$$C^1(\Gamma; G) / C^0(\Gamma; G) \cong \mathrm{Hom}(\pi_1(\Gamma), G) / G$$

where $G$ acts on $\mathrm{Hom}(\pi_1(\Gamma), G)$ by conjugations:

$$\begin{aligned} \mathrm{Hom}(\pi_1(\Gamma), G) \times G &\longrightarrow \mathrm{Hom}(\pi_1(\Gamma), G) \\ (\phi, g) &\longmapsto g^{-1}\phi g \end{aligned}$$

• Tingran Gao, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions." *Discrete & Computational Geometry*, to appear (2019) arXiv:1610.09051

$$C^1(\Gamma; G) / C^0(\Gamma; G) \cong \mathrm{Hom}(\pi_1(\Gamma), G) / G$$

- **LHS**: The *first cohomology set* $\check{H}^1((\Gamma, \mathfrak{U}), \underline{G})$
- **RHS**: The *representation variety* of $\pi_1(\Gamma)$ with value in $G$
- Proof builds upon construction of a *holonomy homomorphism*
- In particular, $\rho \in C^1(\Gamma; G)$ synchronizable
  $\Leftrightarrow [\rho] = [e]$ as equivalence classes in $C^1(\Gamma; G) / C^0(\Gamma; G)$
  $\Leftrightarrow$ the principal $G$-bundle $\mathscr{B}_\rho$ is trivial

# Learning Group Actions by Synchronization

## Problem (Learning Group Actions by Synchronization)

*Denote $\mathscr{X}_K$ for all partitions of $\Gamma$ into $K$ nonempty connected components ($K \leq n$) and*

$$\nu\left(S_i\right) = \inf_{f \in C^0(\Gamma;G)} \sum_{j,k \in S_i} w_{jk} \left\| f_j - \rho_{jk} f_k \right\|^2, \ \mathrm{vol}\left(S_i\right) = \sum_{j \in S_i} d_j,$$

*Solve the optimization problem*

$$\min_{\{S_1,\cdots,S_K\} \in \mathscr{X}_K} \frac{\max\limits_{1 \leq i \leq K} \nu\left(S_i\right)}{\min\limits_{1 \leq i \leq K} \mathrm{vol}\left(S_i\right)} \tag{1}$$

*and output an optimal partition $\{S_1, \cdots, S_K\}$ together with the minimizing $f \in C^0\left(\Gamma; G\right)$.*

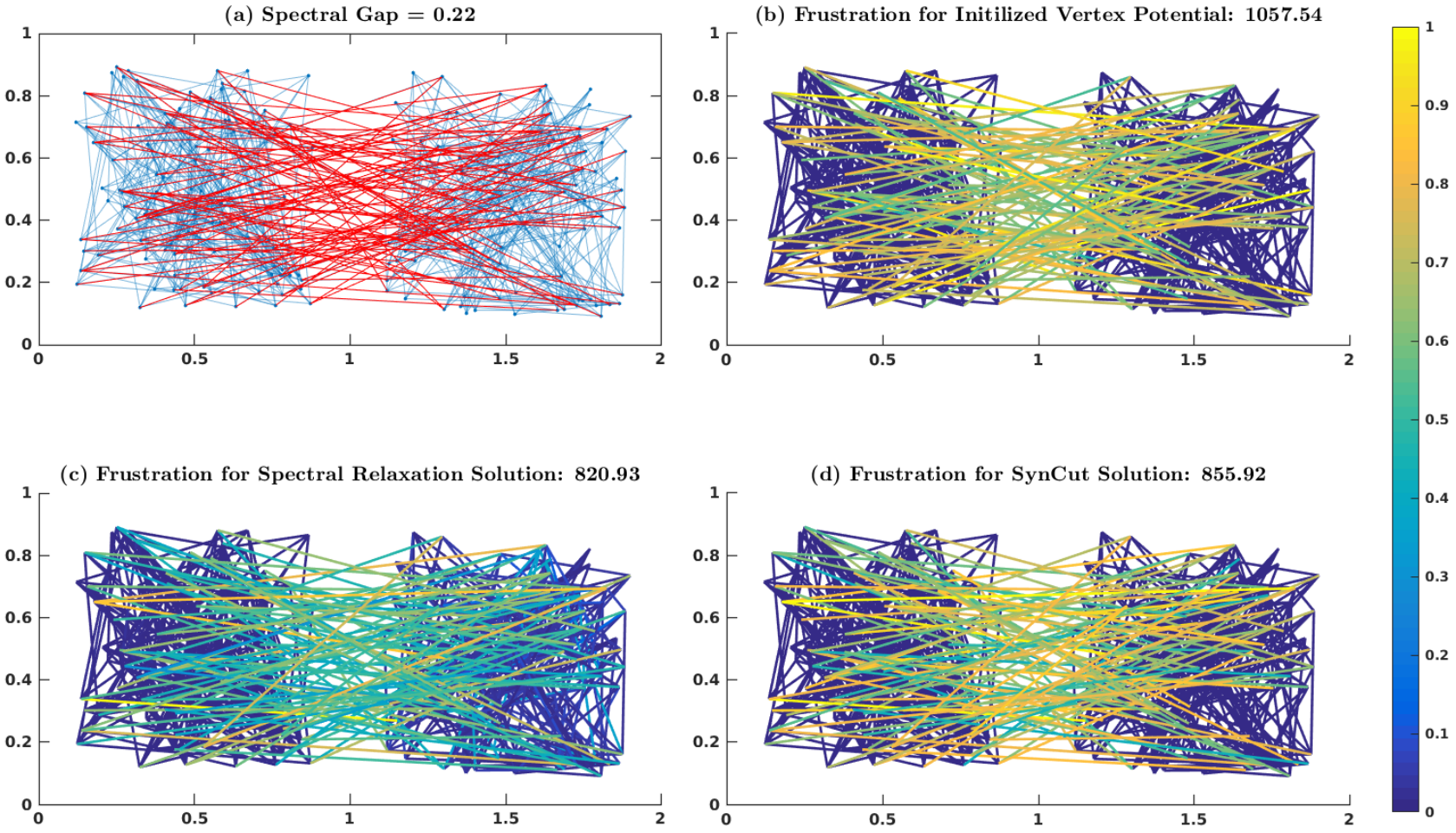
# Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1(\Gamma; G)$, number of partitions $K$
Output: Partitions $\{S_1, \cdots, S_K\}$

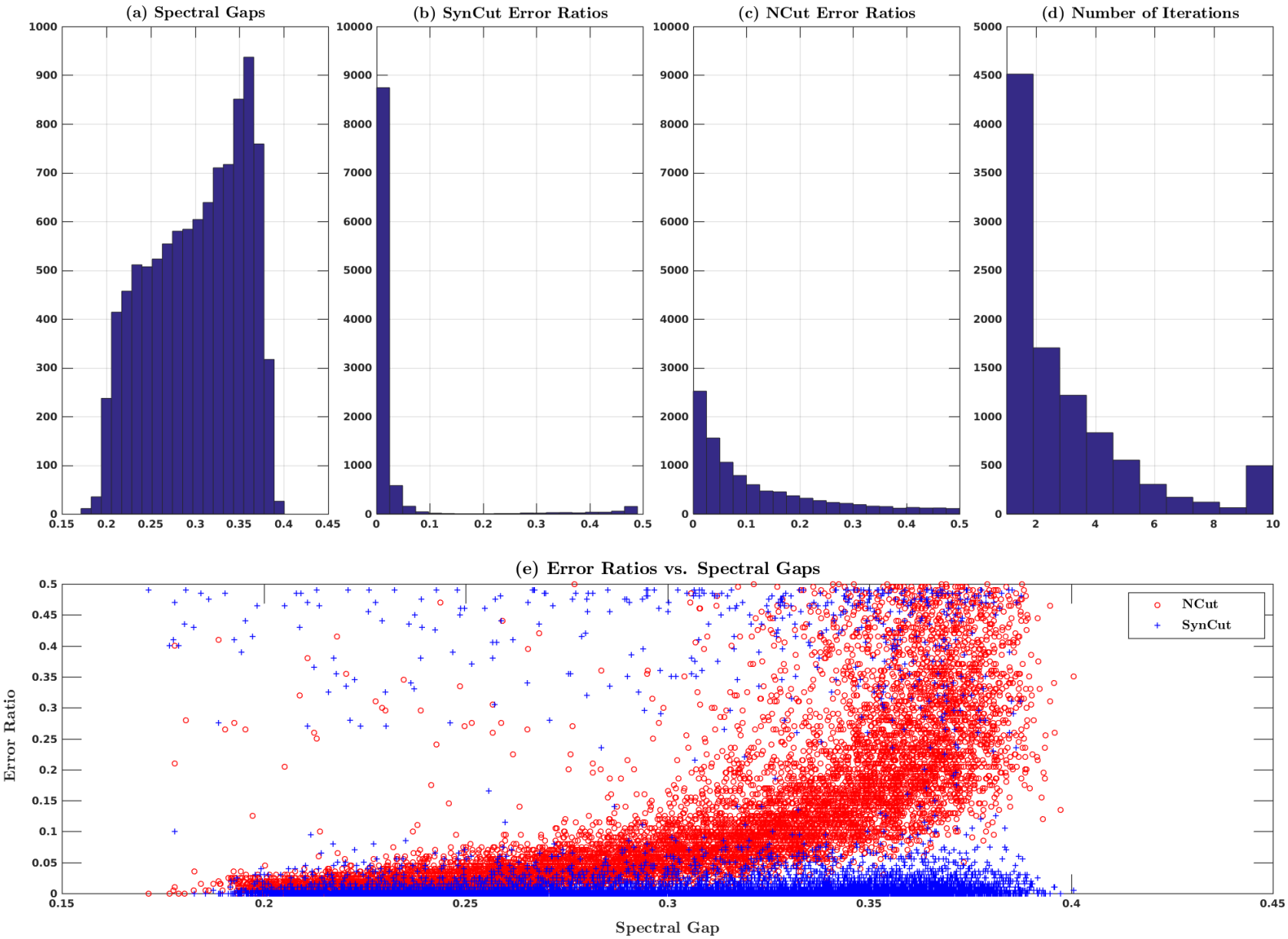
1. Solve synchronization problem over $\Gamma$ for $\rho$, obtain $f \in C^0(\Gamma; G)$
2. Compute $d_{ij} = \exp\left(-w_{ij} \|f_i - \rho_{ij} f_j\|^2\right)$ on all edges $(i, j) \in E$
3. Spectral clustering on weighted graph $(V, E, d)$ to get $\{S_1, \cdots, S_k\}$
4. Solve synchronization problem within each partition $S_j$, "glue up" the local solutions to obtain $f_* \in C^0(\Gamma; G)$
5. $f \leftarrow f_*$, repeat from Step 2


• Tingran Gao, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions." *Discrete & Computational Geometry,* to appear (2019) arXiv:1610.09051
• Chandrajit Bajaj, Tingran Gao, Zihang He, Qixing Huang, and Zhenxiao Liang. "SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions." *Proceedings of the 35th International Conference on Machine Learning,* PMLR 80:324–333, 2018.

# Clustering based on Synchronization Residuals



# Compare with Spectral Clustering



# Outline


Motivation

- Synchronization
- Geometric Model of Fibre Bundles

**Multi-Frequency Phase Synchronization**

- **Phase Synchronization**
- Multi-Frequency Formulation
- Theory

Application to Multi-Frequency Class Averaging


Joint work with **Zhizhen Zhao (UIUC)**

# Phase Synchronization: Synchronization Over $\mathrm{U}(1)$

- Assume the underlying graph $\Gamma$ is complete
- **Ground truth**: phase vector $z = (\exp \iota\theta_1, \ldots, \exp \iota\theta_n)^\top \in \mathbb{C}_1^n$
- **Measurements**: *noisy* pairwise phase offsets $H_{ij} \approx z_i \bar{z}_j$, or in matrix form

$$H := zz^* + \mathrm{noise}$$

- **Goal**: recover the true phase vector $z$ (up to a global multiplicative factor)
- **Spectral Relaxation**: solve for the top eigenvector of $H$, denoted as $\tilde{x}$ (scaled to $\|\tilde{x}\|_2 = \sqrt{n}$), then define $\hat{x} \in \mathbb{C}_1^n$ by

$$\hat{x}_i := \tilde{x}_i / |\tilde{x}_i|$$

# Noise Models

- **Additive Gaussian Noise Model**

$$H = zz^* + \sigma W$$

where $W$ is a standard complex Wigner matrix with i.i.d. standard complex Gaussian entries above the diagonal

- **Random Corruption Model**

$$H_{ij} = \begin{cases} z_i \bar{z}_j & \text{with prob. } r \in [0,1] \\ \mathrm{Unif}\,(\mathrm{U}(1)) & \text{with prob. } 1-r \end{cases}$$

# Optimization Problem Formulation

$$\min_{x \in \mathbb{C}_1^n} \|xx^* - H\|_{\mathrm{F}}^2 \quad \Leftrightarrow \quad \max_{x \in \mathbb{C}_1^n} x^* H x$$

- We may well assume $z = (1, \ldots, 1)^\top$
- Nonconvex problem, can be attacked using convex (spectral or SDP) relaxation
- For additive Gaussian noise, convex relaxation almost recovers the ground truth $z \in \mathbb{C}_1^n$ with high probability for noise level up to $\sigma = O\left(\sqrt{n/\log n}\right)$ **[Zhong & Boumal (2018)]**; non-rigorous statistical-physics-based methods predict the same holds up to $\sigma = O\left(\sqrt{n}\right)$ **[Javanmard et al. (2016)]**
- For random corruption model, **[Singer (2011)]** argues that $|\mathrm{Corr}\,(\tilde{x}, z)| > 1/\sqrt{n}$ with high probability if $r > 1/\sqrt{n}$

# Outline


Motivation

- Synchronization
- Geometric Model of Fibre Bundles

**Multi-Frequency Phase Synchronization**

- Phase Synchronization
- **Multi-Frequency Formulation**
- Theory

Application to Multi-Frequency Class Averaging


Joint work with **Zhizhen Zhao (UIUC)**

# Multi-Frequency Phase Synchronization

- Assume random corruption model
- Instead of solving

$$\max_{x \in \mathbb{C}_1^n} x^* H x$$

  let us solve

$$\boxed{\max_{x \in \mathbb{C}_1^n} \sum_{k=1}^{k_{\max}} (x^k)^* H^{(k)} x^k}$$

  where $x^k := \left(x_1^k, \ldots, x_n^k\right)^\top \in \mathbb{C}_1^n$, and $H^{(k)}$ is the $n$-by-$n$ Hermitian matrix with $H_{ij}^{(k)} := H_{ij}^k$
- Solving this nonconvex optimization by a popular two-stage strategy: (i) produce a good initialization (ii) perform local methods e.g. gradient descent or (generalized) power iteration

• Tingran Gao & Zhizhen Zhao. "Multi-Frequency Phase Synchronization," *The 36th International Conference on Machine Learning* (ICML 2019). arXiv:1901.08235.

# Multi-Frequency Phase Synchronization: Stage 1

- Fix $k_{\max} \geq 1$, build $H^{(2)}, \ldots, H^{(k_{\max})}$ out of $H = H^{(1)}$
- For each $k = 1, \ldots, k_{\max}$, solve the subproblem

$$u^{(k)} := \arg\max_{v \in \mathbb{C}_1^n} \; v^* H^{(k)} v$$

  using any convex relaxation, and set $W^{(k)} := u^{(k)} \left(u^{(k)}\right)^*$
- For all $1 \leq i, j \leq n$, find the "peak location" of the spectrogram

$$\hat{\theta}_{ij} := \arg\max_{\phi \in [0, 2\pi]} \left| \mathrm{Re} \left\{ \sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right\} \right|$$

- Construct Hermitian matrix $\widehat{H} \in \mathbb{C}_1^{n \times n}$ by $\widehat{H}_{ij} = \exp \iota \hat{\theta}_{ij}$, extract the top eigenvector $\tilde{x}$, and perform entrywise normalization to get $\hat{x} \in \mathbb{C}_1^n$

# Multi-Frequency Phase Synchronization: Stage 2

Basically do generalized power iteration, but with a little extra work to ensure the $k_{\max}$ frequency channels "march in formation"

- For iteration counter $t = 0$, set $z^{(k,0)} := \hat{x}^k$ for $k = 1, \ldots, k_{\max}$
- Until convergence ($t = 1, \ldots, T$),
  - $u^{(k,t)} := H^{(k)} z^{(k,t-1)}$ for $k = 1, \ldots, k_{\max}$
  - For each entry $1 \leq i \leq n$, soft-threshold the spectrogram

$$h_i^{(t)} := \mathrm{Re}\left\{ \sum_{k=1}^{k_{\max}} u_i^{(k,t)} e^{\iota k \theta} \right\}$$

  then convert back

$$\hat{u}_i^{(k,t)} := \int_0^{2\pi} \mathrm{SoftThresh}\left( h_i^{(t)}(\theta) \right) \mathrm{d}\theta$$

  and finally get $z_i^{(k,t)} \in \mathbb{C}_1^n$ by entrywise normalization
- Output $\hat{z} := z^{(1,T)}$

# How does it work? Check correlation $|\mathrm{Corr}\,(\hat{z}, z)|$.....

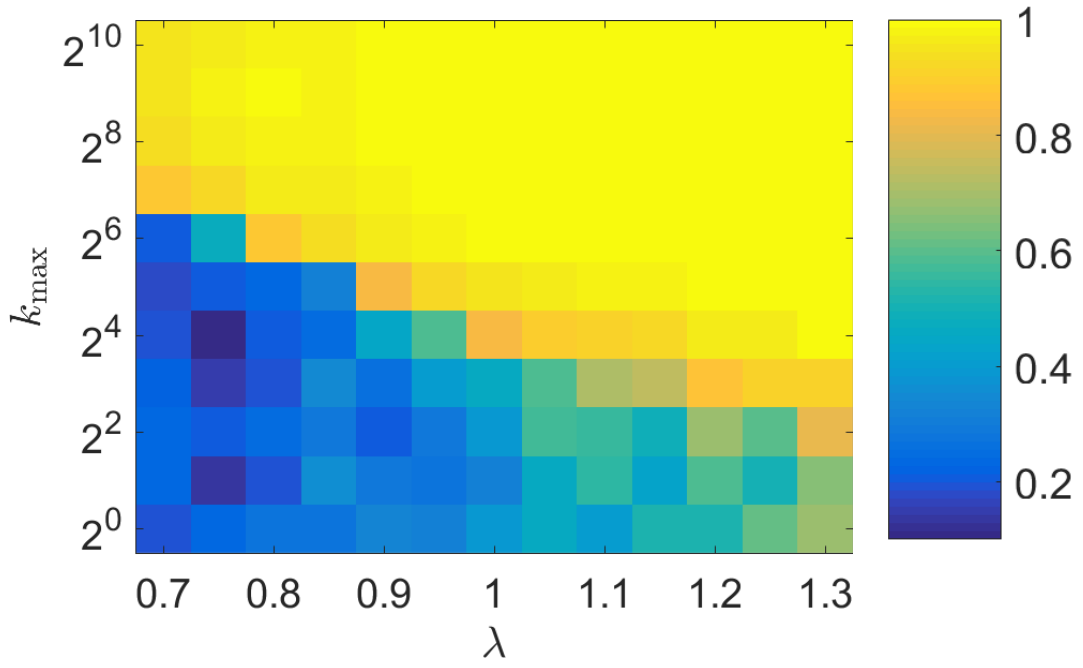



Figure: Random corruption model, $r = \lambda/\sqrt{n}$

- **[Singer (2011)]** only asserts $|\mathrm{Corr}\,(\hat{z}, z)| > 1/\sqrt{n}$ for $\lambda > 1$
- Even for $\lambda < 1$, our method attains $|\mathrm{Corr}\,(\hat{z}, z)| \to 1$ provided that $k_{\max} \gg 1$!

# Outline

Motivation

- Synchronization
- Geometric Model of Fibre Bundles

**Multi-Frequency Phase Synchronization**

- Phase Synchronization
- Multi-Frequency Formulation
- **Theory**

Application to Multi-Frequency Class Averaging

Joint work with **Zhizhen Zhao (UIUC)**

# Focusing on Stage 1

- Fix $k_{\max} \geq 1$, build $H^{(2)}, \ldots, H^{(k_{\max})}$ out of $H = H^{(1)}$
- For each $k = 1, \ldots, k_{\max}$, solve the subproblem

$$u^{(k)} := \arg\max_{v \in \mathbb{C}_1^n} \; v^* H^{(k)} v$$

  using any convex relaxation, and set $W^{(k)} := u^{(k)} \left(u^{(k)}\right)^*$
- For all $1 \leq i, j \leq n$, find the "peak location" of the spectrogram

$$\hat{\theta}_{ij} := \arg\max_{\phi \in [0, 2\pi]} \left| \mathrm{Re} \left\{ \sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right\} \right|$$

- Construct Hermitian matrix $\widehat{H} \in \mathbb{C}_1^{n \times n}$ by $\widehat{H}_{ij} = \exp \iota \hat{\theta}_{ij}$, extract the top eigenvector $\tilde{x}$, and perform entrywise normalization to get $\hat{x} \in \mathbb{C}_1^n$

# Why does it work?

- By a perturbation analysis, we expect

$$\widehat{W}^{(k)} = u^{(k)}(u^{(k)})^* = z^k(z^k)^* + E^{(k)} \approx z^k(z^k)^*$$

from solving each subproblem

$$u^{(k)} := \arg\max_{v \in \mathbb{C}_1^n} \; v^* H^{(k)} v$$

- The peak finding step is expected to ensure

$$\begin{aligned}\hat{\theta}_{ij} &= \arg\max_{\phi \in [0,2\pi]} \left| \mathrm{Re}\left\{ \sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right\} \right| \\ &\approx \arg\max_{\phi \in [0,2\pi]} \left| \mathrm{Re}\left\{ \sum_{k=1}^{k_{\max}} z_i^k \bar{z}_j^k e^{-\iota k \phi} \right\} \right| = \theta_i - \theta_j\end{aligned}$$

provided that $E^{(k)}$ does not "perturb away" the maximum!

# Landscape Analysis of the Dirichlet Kernel

$$2\mathrm{Re}\left\{\sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k\phi}\right\} = \sum_{k=-k_{\max}}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k\phi}$$

$$= \sum_{k=-k_{\max}}^{k_{\max}} e^{\iota k\left(\theta_i-\theta_j\right)} e^{-\iota k\phi} + \sum_{k=-k_{\max}}^{k_{\max}} E_{ij}^{(k)} e^{-\iota k\phi}$$

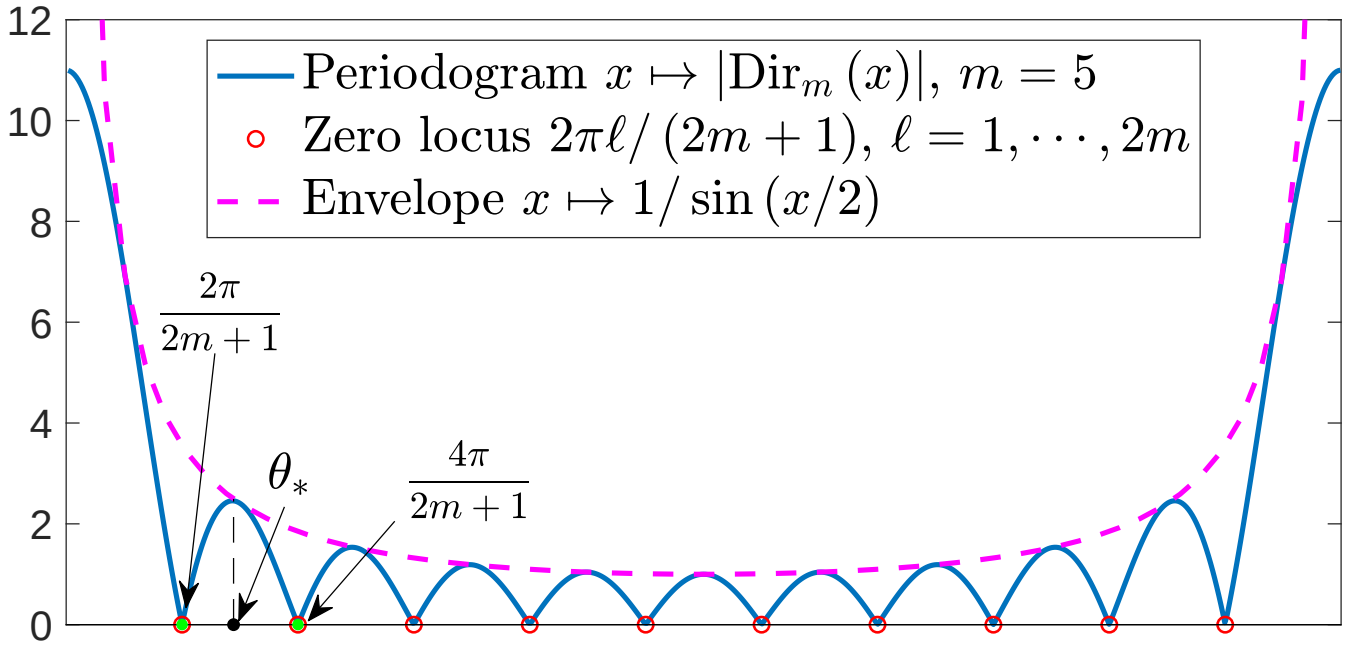
# Landscape Analysis of the Dirichlet Kernel

$$
\begin{aligned}
2\mathrm{Re}\left\{\sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k\phi}\right\} &= \sum_{k=-k_{\max}}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k\phi} \\
&= \sum_{k=-k_{\max}}^{k_{\max}} e^{\iota k\left(\theta_i-\theta_j\right)} e^{-\iota k\phi} + \sum_{k=-k_{\max}}^{k_{\max}} E_{ij}^{(k)} e^{-\iota k\phi} \\
&= \frac{\sin\left[\left(k_{\max}+\frac{1}{2}\right)\left(\theta_i-\theta_j-\phi\right)\right]}{\sin\left[\frac{1}{2}\left(\theta_i-\theta_j-\phi\right)\right]} + \sum_{k=-k_{\max}}^{k_{\max}} E_{ij}^{(k)} e^{-\iota k\phi} \\
&=: \mathrm{Dir}_{k_{\max}}\left(\theta_i-\theta_j-\phi\right) + R_{k_{\max}}\left(\phi\right)
\end{aligned}
$$

# Landscape Analysis of the Dirichlet Kernel

**Key Observation:** $\left|\hat{\theta}_{ij} - (\theta_i - \theta_j)\right| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

$$2k_{\max} + 1 - R_{k_{\max}}(\phi) > \frac{1}{\sin(\theta_*/2)} + R_{k_{\max}}(\phi)$$



# Landscape Analysis of the Dirichlet Kernel

**Key Observation:** $\left|\hat{\theta}_{ij} - (\theta_i - \theta_j)\right| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

$$2k_{\max} + 1 - R_{k_{\max}}(\phi) > \frac{1}{\sin(\theta_*/2)} + R_{k_{\max}}(\phi)$$

which is satisfied if

$$2k_{\max} + 1 - R_{k_{\max}}(\phi) > \frac{1}{\sin\left(\frac{\pi}{2k_{\max}+1}\right)} + R_{k_{\max}}(\phi)$$

$$\Leftarrow \frac{1}{k_{\max}} \sum_{k=-k_{\max}}^{k_{\max}} E_{ij}^{(k)} e^{-\iota k\phi} = \frac{1}{k_{\max}} R_{k_{\max}}(\phi) < 1 - \frac{1}{2k_{\max}\sin\left(\frac{\pi}{2k_{\max}+1}\right)}$$

**We just need a uniform upper bound for the $E_{ij}^{(k)}$'s!**

# Eigenvector perturbation analysis in $\ell_\infty$-norm!

- Standard perturbation bounds use $\ell_2$-norm (e.g. Davis–Kahan), but most application scenarios of spectral methods require bounding the $\ell_\infty$-norm
- Active research area in recent years, e.g. **[Eldridge et al. (2017)]; [Abbe et al. (2017)]; [Fan et al. (2018)]; [Zhong & Boumal (2018)]**
- Sharpest results to date use a "leave-one-out" trick popularized by statisticians

**Lemma (Gao & Zhao 2019).** For any $0 < \epsilon \leq 2$, with probability at least $1 - O\left(n^{-(2+\epsilon)}\right)$, there exists absolute constant $C > 0$ such that

$$\left\| E_{ij}^{(k)} \right\|_\infty \leq C \sqrt{\frac{\log n}{n}}.$$

• Tingran Gao & Zhizhen Zhao. "Multi-Frequency Phase Synchronization," *The 36th International Conference on Machine Learning* (ICML 2019). arXiv:1901.08235.

# Putting Everything Together

**Theorem (Gao & Zhao 2019).** With all (mild) assumptions satisfied, with high probability the multi-frequency phase synchronization algorithm produces an estimate $\hat{z}$ satisfying

$$\mathrm{Corr}\left(\hat{z}, z\right) \geq 1 - \frac{C'}{k_{\max}^2}$$

provided that

$$k_{\max} > \max\left\{5, \frac{1}{\sqrt{2}\,\pi\left(1 - 4C_2\sigma\sqrt{\log n/n}\right) - 2}\right\}.$$

In particular, $\mathrm{Corr}\left(\hat{z}, z\right) \to 1$ as $k_{\max} \to \infty$.

• Tingran Gao & Zhizhen Zhao. "Multi-Frequency Phase Synchronization," *The 36th International Conference on Machine Learning* (ICML 2019). arXiv:1901.08235.

# Outline

Motivation

- Synchronization
- Geometric Model of Fibre Bundles

Multi-Frequency Phase Synchronization

- Phase Synchronization
- Multi-Frequency Formulation
- Theory

**Application to Multi-Frequency Class Averaging**

Joint work with **Zhizhen Zhao (UIUC)**

# Class Averaging and the Intrinsic Model

- **Goal:** Classify cryo-EM images $I_1, I_2, \ldots$ according to their viewing directions
- **Procedure:**
  - Compute the *rotation-invariant distance* between all pairs of images $d_{\mathrm{RID}}(I_i, I_j) := \min_{\alpha \in [0,2\pi]} \|I_i - e^{\iota\alpha} I_j\|_{\mathrm{F}}$, and denote $\alpha_{ij}$ for the optimal alignment angle
  - Fix threshold $\epsilon > 0$ and define Hermitian $W \in \mathbb{C}^{n \times n}$ by

$$W_{ij} := \begin{cases} \exp(\iota\alpha_{ij}) & \text{if } d_{\mathrm{RID}}(I_i, I_j) < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

  - Solve for the top three eigenvectors of $W$, which embeds $I_1, I_2, \ldots$ into $\mathbb{R}^3$
  - Use correlation in the embedded $\mathbb{R}^3$ space to determine the closeness between viewing directions

# Multi-Frequency Class Averaging

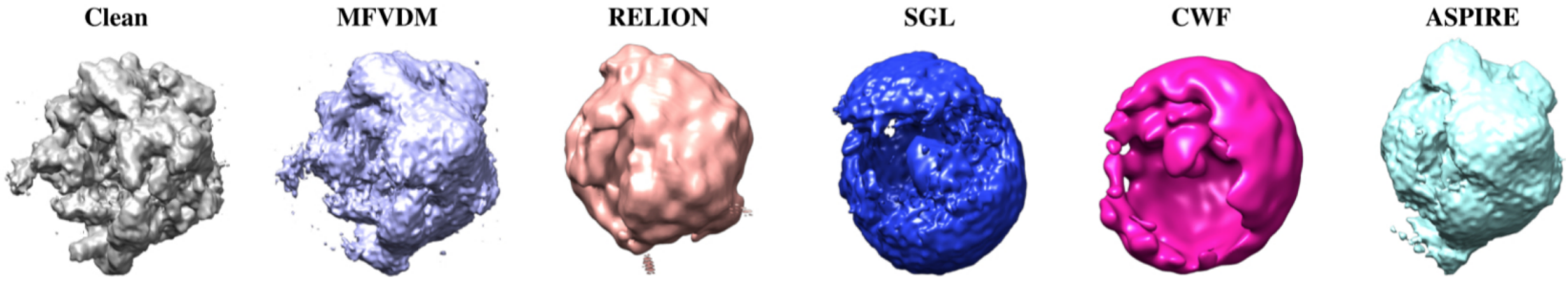
- Idea directly descended from the multi-frequency phase synchronization, but the underlying fibre bundle is non-trivial!
- Fix threshold $\epsilon > 0$ and define for each $k = 1, \ldots, k_{\max}$ Hermitian $W^{(k)} \in \mathbb{C}^{n \times n}$ by

$$W_{ij}^{(k)} := \begin{cases} \exp\left(\iota k \alpha_{ij}\right) & \text{if } d_{\mathrm{RID}}^{(k)}\left(I_i, I_j\right) < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

- Solve for the top $2k+1$ eigenvectors of $W^{(k)}$, which embeds $I_1, I_2, \ldots$ into $\mathbb{R}^{2k+1}$
- Use all correlations in the embedded $\mathbb{R}^{2k+1}$ $(k = 1, \ldots, k_{\max})$ spaces to determine the closeness between viewing directions

• Yifeng Fan & Zhizhen Zhao. "Cryo-Electron Microscopy Image Analysis Using Multi-Frequency Vector Diffusion Maps," *preprint*. arXiv:1904.07772.

# Multi-Frequency Information Improves Class Averaging

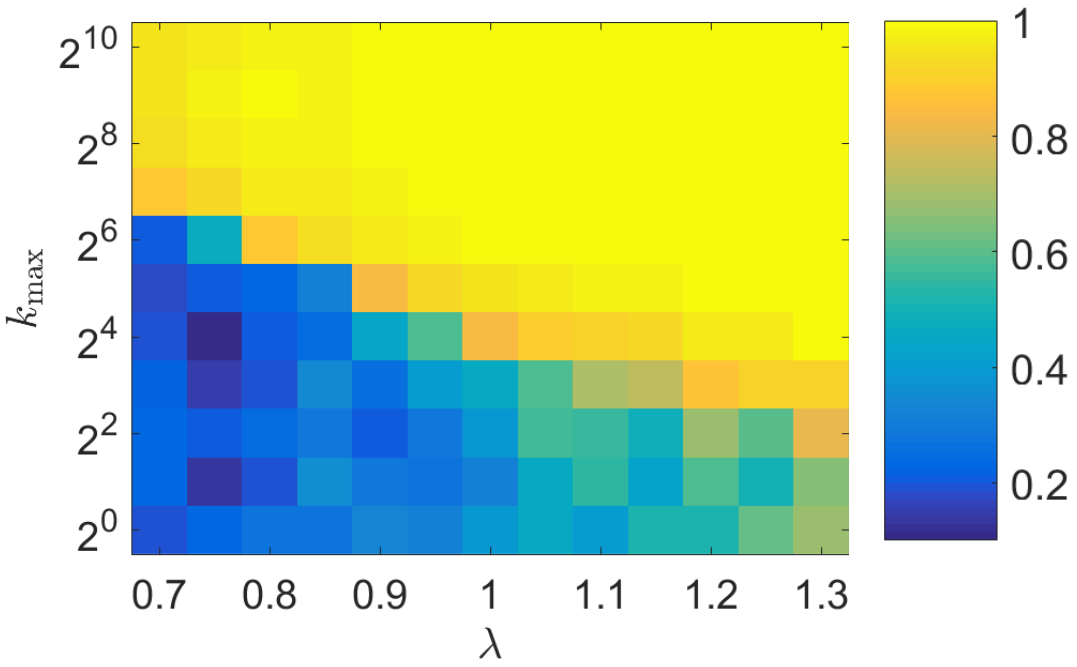



• Yifeng Fan & Zhizhen Zhao. "Cryo-Electron Microscopy Image Analysis Using Multi-Frequency Vector Diffusion Maps," *preprint*. arXiv:1904.07772.

# Open Problems

- Landscape and convergence analysis for the Stage 2 algorithm?
- Algorithm works for $\mathrm{SO}(3)$ seamlessly, but theory?
- Possible extension to other synchronization and multireference alignment problems over compact/noncompact Lie groups?
- Mult-frequency vector diffusion maps?
- A learning paradigm on sheaves?

# Thank You!





• Tingran Gao & Zhizhen Zhao. "Multi-Frequency Phase Synchronization," *The 36th International Conference on Machine Learning* (ICML 2019). arXiv:1901.08235.